

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.

MENU

SEARCH

INDEX

DETAIL

1/1



JAPANESE PATENT OFFICE

PATENT ABSTRACTS OF JAPAN

(11)Publication number: 08087508

(43)Date of publication of application: 02.04.1996

(51)IntCl.

G06F 17/30

(21)Application number: 06220678

(71)Applicant:

OLYMPUS OPTICAL CO LTD

(22)Date of filing: 14.09.1994

(72)Inventor:

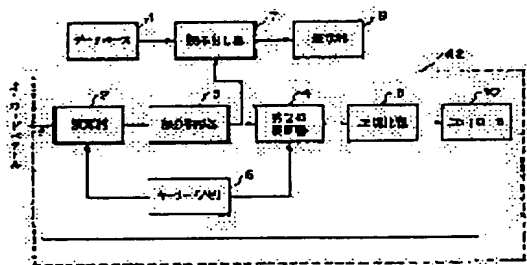
OOYAMA NAGAAKI
WADA TOSHIAKI

(54) INFORMATION RETRIEVING DEVICE

(57)Abstract:

PURPOSE: To provide an information retrieving device which can suitably correct a pointed retrieving key word for retrieval in response to the prescribed inter-key word correlation and then takes the desired information out of a data base.

CONSTITUTION: An information retrieving device includes a computing element 2 which calculates the inner product between the retrieving key word matrix R_{in} from a key word memory 6 and the retrieving vector that is corrected in response to the correlation, a partial linear unit 3 which converts the calculation result of the element 2 into a partial linear shape, a 2nd computing element 4 which multiplies the output vector of the unit 3 by a key word matrix X , and a normalizer unit 5 which normalizes the elements of the product result into 0 and 1. Furthermore a controller 10 is added to compare the output vector of the unit 5 that is not fed back once to the element 2 with the output vector of the unit 3, together with a reader 7 which reads the desired data out of a data base 1 based on the address corresponding to the corrected output vector of the unit 3, and a display device 8 which shows the desired data.



LEGAL STATUS

[Date of request for examination]
[Date of sending the examiner's decision of rejection]
[Kind of final disposal of application other than the
examiner's decision of rejection or application converted
registration]
[Date of final disposal for application]
[Patent number]
[Date of registration]
[Number of appeal against examiner's decision of rejection]
[Date of requesting appeal against examiner's decision of
rejection]
[Date of extinction of right]

Copyright (C); 1998 Japanese Patent Office

[MENU](#)

[SEARCH](#)

[INDEX](#)

[DETAIL](#)

(12) 公開特許公報 (A)

KDD - 32

(11)特許出願公開番号

特開平8-87508

(43)公開日 平成8年(1996)4月2日

(51) Int.Cl.⁸

G O 6 F · 17/30

識別記号

庁内整理番号

FI

技術表示箇所

9194-5L

G O 6 F 15/ 403

340 C

審査請求 未請求 請求項の数15 O L (全 12 頁)

(21)出願番号

特願平6-220678

(22) 出願日

平成6年(1994)9月14日

(71)出願人 000000376

オリンパス光学工業株式会社

東京都渋谷区幡ヶ谷2丁目43番2号

(72)発明者 大山 永昭

神奈川県横浜市緑区長津田町4259 東京工
業大学内

(72)発明者 和田 利昭

東京都渋谷区幡ヶ谷2丁目43番2号 オリ
ンパス光学工業株式会社内

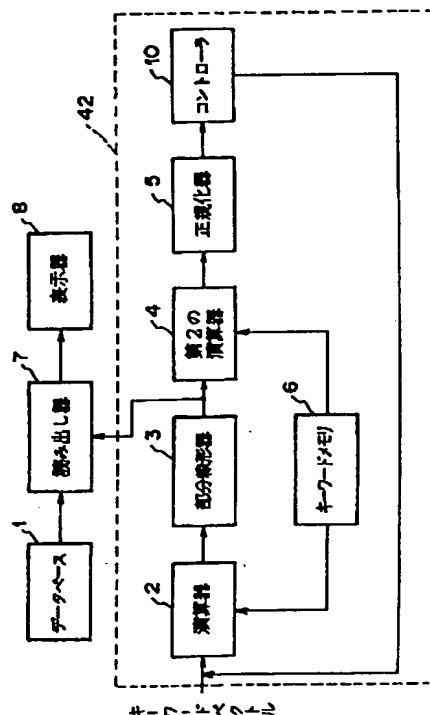
(74) 代理人 弁理士 鈴江 武彦

(54) 【発明の名称】 情報検索装置

(57) 【要約】

【目的】本発明は、指示された検索キーワードを予め定められたキーワード間の相関度に応じて検索に好適するように修正し、データベースから所望する情報を取り出す情報検索装置を提供することを目的とする。

【構成】本発明は、キーワードメモリ6からの索引キーワード行列と相関度に応じて修正された検索ベクトルとを内積する演算器2と、その結果を部分的線形に変換する部分線形器3と、部分線形器3の出力ベクトルとキーワード行列Xとを積する第2演算器4と、積結果の各要素に対して0, 1に正規化する正規化器5と、演算器2に1回フィードバックする前の正規化器5の出力ベクトルと比較するコントローラ10と、修正された部分線形器3の出力ベクトルの位置の対応するアドレスに基づき、データベース1から所望データを読み出す読み出し器7と、その所望データを表示する表示器8とで構成される情報検索装置である。



【特許請求の範囲】

【請求項1】 データベースを検索する情報検索装置において、

所定の検索キーワードを入力し、数値ベクトル化する検索キーワードベクトル化手段と、

前記数値ベクトル化された検索キーワードベクトルをキーワード間の相関度に応じて修正する検索キーワードベクトル修正手段と、

前記修正された検索キーワードベクトルに基づいて検索候補を出力する検索手段と、

前記検索候補から、所望する情報を選択する選択手段と、

前記キーワード間の相関度を修正するキーワード間相関度修正手段と、を具備することを特徴とする情報検索装置。

【請求項2】 前記検索キーワードベクトル修正手段は、キーワードのカテゴリ毎に修正することを特徴とする請求項1記載の情報検索装置。

【請求項3】 前記検索キーワードベクトル化手段は、キーワードのカテゴリ毎に1つのみ選択したキーワードを1、残り全てを0とするベクトルに変換することを特徴とする請求項1及び2記載の情報検索装置。

【請求項4】 前記キーワード間相関度修正手段が、前記検索キーワードを表す数値ベクトルと、前記選択した情報に付されている索引キーワードを表す数値ベクトルの複数の組により算出する手段により構成されていることを特徴とする請求項1乃至3記載の情報検索装置。

【請求項5】 データベースを検索する情報検索装置において、

所定の検索キーワードに基づいて、数値ベクトル化する索引キーワードベクトル化手段と、

前記数値ベクトル化された検索キーワードベクトルを、キーワードのカテゴリ毎にキーワード間の相関度に応じて修正する検索キーワードベクトル修正手段と、

前記修正された検索キーワードベクトルを入力し、検索候補を出力する検索手段と、

前記検索候補から、所望の情報を選択する選択手段と、

選択された所望情報に数値ベクトルの形で付されている索引キーワードベクトルを前記検索キーワードベクトルと共に記憶するキーワードベクトル記憶手段と、

前記キーワードベクトル記憶手段に記憶されている複数の検索及び索引キーワードベクトルの組から前記キーワード間の相関度を算出するキーワード相関度算出手段と、

前記キーワード相関度算出手段により算出された値により前記キーワード間の相関度を修正するキーワード間相関度修正手段と、を具備することを特徴とする情報検索装置。

【請求項6】 前記検索器は検索候補をベクトルとして出力し、そのベクトルから検索候補の情報の存在するアドレスを指定する手段を、さらに具備することを特徴と

する請求項5記載の情報検索装置。

【請求項7】 データベースを検索する情報検索装置において、

利用者を特定する手段と、

利用者毎に前記キーワード相関テーブルと、を具備することを特徴とする請求項1乃至6記載の情報検索装置。

【請求項8】 データベースを検索する情報検索装置において、

検索時に利用者が自分専用のキーワード相関テーブルを与える手段を、さらに具備することを特徴とする請求項7記載の情報検索装置。

【請求項9】 データベースを検索する情報検索装置において、

前記利用者個人の専用のキーワード相関テーブルを情報検索装置の外部の記憶媒体に保持していることを特徴とする請求項7及び8記載の情報検索装置。

【請求項10】 データベースを検索する情報検索装置において、

所定の検索キーワードデータを入力し数値ベクトル化する索引キーワードベクトル化手段と、

その検索キーワードの確信度を入力する手段と、

前記数値ベクトル化された検索キーワードベクトルを前記確信度に応じて修正する検索キーワードベクトル修正手段と、

前記修正された検索キーワードベクトルに基づいて、検索候補を出力する検索手段と、

前記検索候補から、所望の情報を選択する選択手段と、

具備することを特徴とする情報検索装置。

【請求項11】 前記検索キーワードベクトル修正手段は、検索キーワードのカテゴリ毎に修正し、

確信度が1の場合には未修正とし、

確信度が0の場合にはそのキーワードが属するカテゴリの全てのキーワードに同じ数値となるように修正し、

1と0の間の場合には選択したキーワードについては値を小さくし、それ以外のキーワードについては値を大きくすることを特徴とする請求項10記載の情報検索装置。

【請求項12】 前記検索手段が連想過程を1乃至複数回繰り返しながら行うことにより検索を実行することを特徴とする請求項1乃至11記載の情報検索装置。

【請求項13】 前記検索手段がニューラルネットワークであることを特徴とする請求項12記載の情報検索装置。

【請求項14】 前記検索手段が前記検索キーワードベクトル修正手段を含んで処理することを特徴とする請求項1乃至13記載の情報検索装置。

【請求項15】 前記検索手段が単一の演算器により構成されることを特徴とする請求項1乃至14記載の情報検索装置。

【請求項16】 前記検索手段が単一の演算器により構成されることを特徴とする請求項1乃至14記載の情報検索装置。

【請求項17】 前記検索手段が単一の演算器により構成されることを特徴とする請求項1乃至14記載の情報検索装置。

【請求項18】 前記検索手段が単一の演算器により構成されることを特徴とする請求項1乃至14記載の情報検索装置。

【発明の詳細な説明】

10

20

30

40

50

【0001】

【産業上の利用分野】本発明は、種々の情報を多数記憶したデータベースから所望する情報を取り出す情報検索装置に関する。

【0002】

【従来の技術】一般に、データベースに記憶された多数の情報の中から所望する情報を取り出す場合には、予めそれぞれの情報に付与されているキーワードを手掛かりに検索を行なっている。この時、データベースの利用者は、複数のキーワードの論理式を入力し、キーワードマ

atchingすれば、必要な情報を引き出すことができる。【0003】実際には、データベースに好適するキーワードが設定できない場合もあり、これに対応するため、キーワードの一部として、いかなる文字にもヒットするワイルドカードが用意されている。例えば、公知のUNIXシステムやMS-DOSシステムでは、“*”は任意の文字数の任意の文字にヒットする。この場合、“東京*”で検索すると、“東京都”や“東京大学”など、“東京～”の単語にヒットする。また、キーワードを体系化したシソーラスを備え、入力したキーワードの下位概念に対応するキーワードを追加して検索することが可能な検索装置もある。

【0004】さらには、データベース化された情報に付された各索引キーワードと、ユーザーが入力した検索キーワードを数値ベクトルで表現し、検索キーワードベクトルと各索引キーワードベクトル間の距離を利用した検索方法が考案されている。

【0005】例えば、特開平2-224068号公報に開示される「情報検索システム」によれば、キーワード間の相関の度合をデータベースに登録されている文献情報内の各キーワードの出現頻度に基づいて、

【0006】

【数1】

$$K_{ij} = \frac{F(i, j)}{F(j)}$$

で求め、この相関度により検索キーワードベクトルを変更して、より高機能な文書検索を実現している。ここで、 K_{ij} はキーワード*i*と*j*の相関度、 $F(j)$ はキーワード*j*の記憶されている文書中の出現頻度、 $F(i, j)$ はキーワード*i*と*j*が同時に、その文書中に出現する頻度である。

【0007】

【発明が解決しようとする課題】しかし、前述したUNIXシステムやMS-DOSシステムでは、たとえデータベースに情報を登録した本人が検索する場合であっても、人間の記憶は時間とともに曖昧になってしまう。

【0008】その結果、従来のような2値論理に基づく検索装置では、検索時に利用者が間違ったキーワードを設定した場合には間違った検索が行なわれ、利用者が求めるデータを検索するために数多くの候補を読み出さな

ければならなかったり、検索が全く不可能だったりする場合がある。

【0009】また上記公報に記載される情報検索システムでは、データベースに登録されている情報がキーワードとして直接利用できる文書の場合でなければ使えないので、写真や描画等の画像データベースに応用することはできない。

【0010】そこで本発明は、指示された検索キーワードを予め定められたキーワード間の相関度に応じて検索に好適するように修正し、さらにその相関度も修正するようにしてデータベースから所望する情報を取り出す情報検索装置を提供することを目的とする。

【0011】

【課題を解決するための手段】本発明は上記目的を達成するために、第1に、データベースを検索する情報検索装置において、所定の検索キーワードを入力し、数値ベクトル化する検索キーワードベクトル化手段と、前記数値ベクトル化された検索キーワードベクトルをキーワード間の相関度に応じて修正する検索キーワードベクトル修正手段と、前記修正された検索キーワードベクトルに基づいて検索候補を出力する検索手段と、前記検索候補から、所望する情報を選択する選択手段と、前記キーワード間の相関度を修正するキーワード間相関度修正手段とで構成された情報検索装置を提供する。

【0012】第2に、データベースを検索する情報検索装置において、所定の検索キーワードを入力し数値ベクトル化する検索キーワードベクトル化手段と、前記数値ベクトル化された検索キーワードベクトルを、キーワードのカテゴリ毎にキーワード間の相関度に応じて修正する検索キーワードベクトル修正手段と、前記修正された検索キーワードベクトルに基づいて、検索候補を出力する検索手段と、前記検索候補から、所望の情報を選択する選択手段と、選択された所望情報に数値ベクトルの形で付されている索引キーワードベクトルを前記検索キーワードベクトルと共に記憶するキーワードベクトル記憶手段と、前記キーワードベクトル記憶手段に記憶されている複数の検索及び索引キーワードベクトルの組から前記キーワード間の相関度を算出するキーワード相関度算出手段と、前記キーワード相関度算出手段により算出された値により前記キーワード間の相関度を修正するキーワード間相関度修正手段とで構成された情報検索装置を提供する。

【0013】第3に、データベースを検索する情報検索装置において、所定の検索キーワードデータを入力し数値ベクトル化する検索キーワードベクトル化手段と、その検索キーワードの確信度を入力する手段と、前記数値ベクトル化された検索キーワードベクトルを前記確信度に応じて修正する検索キーワードベクトル修正手段と、前記修正された検索キーワードベクトルを入力し、検索候補を出力する検索手段と、前記検索候補から、所望の

情報を選択する選択手段とで構成された情報検索装置を提供する。

【0014】

【作用】以上のような構成の情報検索装置は、まず、利用者が入力した検索キーワードと、利用者が正しい検索結果として読出された情報に付加されていた索引キーワードを用いて各検索キーワードの相関テーブルを作成し、この相関テーブルで検索キーワードが修正される。

【0015】そして、検索キーワードをキーワード間の相関度に応じて修正する機能と、キーワード間の相関度を実際の検索結果から求める手段を有するので、不十分な検索キーワードからも効果的に検索される。また、個人個人の概念の違いに対応するため、利用者を特定し、その利用者用のキーワードの相関テーブルを作成し、この相関テーブルを用いて検索する。

【0016】また、各検索時に検索キーワードと共にそのキーワードに対する確信度を入力し、確信度が異なることによるキーワード間の相関度の変化を考慮して検索する。従って、確信度によりキーワード相関度テーブルが修正され、利用者の記憶の状況が検索に反映される。

【0017】

【実施例】以下、図面を参照して本発明の実施例を詳細に説明する。図8には、本発明による第1実施例としての情報検索装置を含むファイル検索システムの概略的な構成を示し説明する。本実施例では、データベースに画像データがファイリングされているものとして説明するが、キーワードを付与することが可能な情報であれば、ファイリングされているデータはどのような形態のものでもよい。

【0018】図8に示すシステムにおいて、データベース1と、所定検索キーワードを入力するためのキーボード等からなるキーワード入力器40と、キーワード入力器40が出力する検索キーワードベクトルと、その検索キーワードに対して正しく読出された情報に付されていた索引キーワードベクトルの組を記憶する蓄積器44と、前記蓄積器44に記憶された検索キーワードベクトルと、索引キーワードベクトルの組を基にキーワード間の相関を計算する演算器45と、検索キーワードベクトルを各カテゴリーごとにキーワード間の相関度を記憶するキーワード相関テーブル46と、キーワード相関テーブル46に記憶されているキーワード間の相関度に応じて検索キーワードを修正する変換器41と、修正された検索キーワードベクトルに基づき所定の出力ベクトルを検索する検索器42と、前記検索器42の出力ベクトルの各要素の値が大きい順に所定数選択する選択器43と、選択器43に選択された出力ベクトル（アドレス）に従い、データベース1から所望のデータを読出す読み出し器7と、読み出されたデータを表示する表示部8とで構成される。

【0019】このように構成されたシステムにおいて、

データベース利用者がキーワード入力器40を操作して、予め定められているキーワード中からいくつかを選択して入力する。ここで、全キーワードは所定のカテゴリーに分類されており、利用者は各カテゴリーから1個のキーワードを選択するものとする。各カテゴリーは選択したキーワードに対応する値のみ“1”で、同じカテゴリーに属する他のキーワードに対応する値は全て“0”となるようにコード化される。

【0020】例えば、図2(a)に示すように、天候をカテゴリーの1つとし、このカテゴリーに「晴れ」、「曇り」、「雨」、「雪」の4個のキーワードが属しているものとする。利用者が「晴れ」のキーワードを選択すれば、「晴れ」に対応する数値が“1”で他の数値は全て“0”となる。「曇り」を選択した場合には、「曇り」の数値が“1”で他の数値は“0”となる。「雨」や「雪」の場合も同様に「雨」や「雪」の数値のみ“1”で他の数値は“0”となる。このように、選ばれた1つの天候の数値のみ“1”で、他の数値は“0”となる4要素の数値ベクトルで表される。

【0021】また図2(b)には、データベースに記憶されたデータに付与されたそれぞれの索引キーワードのベクトルの一例を示している。検索ベクトルとしては、全てのカテゴリーのベクトルにより構成される図2(c)のような1つのベクトルである。

【0022】前記変換器41はキーワード入力器40が出力する検索キーワードベクトルを各カテゴリーごとにキーワード相関テーブル46に記憶されているキーワード間の相関度に応じて修正する。

【0023】各カテゴリーに属するキーワードの相関度は、図2(d)に示すような行列の形でキーワード相関テーブル46に記憶されている。例えば、天候カテゴリーに関するキーワード部分について述べると、変換器41は検索キーワードベクトルの天候のカテゴリーに関する部分ベクトルに、この図2(d)のような相関行列を掛けることにより修正する。全てのカテゴリーが修正されると、修正された検索キーワードベクトルは検索器42に入力する。

【0024】次に、図1には、本発明の情報検索装置の具体的な構成例となる前記検索器42の構成を示す。まずキーワードメモリ6には、データベース1に記憶している各情報に付与されている索引キーワードがベクトルの集合として記憶されている。この索引キーワードは、前述した検索キーワードと同様な構造を有した、各成分“0”又は“1”のベクトルである。

【0025】そして演算器2では、このキーワードメモリ6に記憶されている全ての索引キーワードベクトルより構成される索引キーワード行列と前記変換器41で修正された検索ベクトルとの内積が計算される。すなわち、前記検索ベクトルを p 、記憶されている n 個の索引キーワードベクトルを x_i 、($i=1, 2, \dots, n$)と

7

し、これら記憶ベクトルを列ベクトルとして構成される索引キーワード行列を $X = [x_1, x_2, \dots, x_n]$ 、その転置行列を X^T とすれば、 $X^T p$ が計算される。ここで、 $X^T p$ は各々の索引ベクトル x_i と p の内積 $x_i^T p$ を要素とするベクトルとなる。

【0026】これは、 $X^T p$ の計算結果のそれぞれの要素について、検索キーワードベクトルと各索引キーワードベクトルとの近さを表す。前記計算結果のベクトル $X^T p$ の各要素に対して、部分線形器3で図3に示すような部分的に線形となる変換 g を行う。この変換処理では、入力値が a 未満の場合には“0”を出力し、 a 以上の場合には入力値がそのまま出力する。この処理により前記検索ベクトルと相関の低い成分は“0”となり、以降の処理に関与しなくなる。

【0027】前記部分線形器3の出力ベクトル $g(X^T p)$ は、第2の演算器4において、前記キーワードメモリ6に記憶されているキーワード行列 X との積が計算される。 $X g(X^T p)$ は、検索キーワードベクトルと各々の索引キーワードベクトルの近さを重みとしてこれらの記憶ベクトルを線形結合したものである。このベクトルは、異なるキーワード間の相関の度合に応じて検索キーワードとして与えられた要素以外についても“0”でない値を持つ。

【0028】この $X g(X^T p)$ の結果を正規化器5で所定の正規化した後、フィードバックし、新たな検索ベクトルとして再び入力することにより利用者がはじめに入力した検索ベクトル p で設定したキーワードのみでなく、このキーワードと相関のあるキーワードについても検索が実行されることになる。このような処理を定常状態に達するまで繰り返す。

【0029】そして正規化器5では、 $X g(X^T p)$ の結果の各要素に対して、要素の値が“0”以下の場合には“0”、また“0”を越える場合は“1”となるように正規化する。この結果、正規化器5の出力は要素が“0”と“1”のベクトルとなる。この正規化器5の出力ベクトルはフィードバックして前記演算器2に入力し、上記した各処理を繰り返す。

【0030】図4に示すように、前記部分線形器3の出力ベクトルの各要素の該ベクトル中の位置は、それぞれの記憶ベクトルに対応しているため、この位置をデータベース中のアドレスに対応づけることにより情報の検索ができる。すなわち、定常状態に達した後、部分線形器3の出力ベクトルの要素の値が高い順番に情報を読み出せばよい。

【0031】そして、コントローラ10は前記正規化器5の出力ベクトルを入力し、1回だけフィードバックする前の正規化器5の出力ベクトルと比較する。今回、出力したベクトルが1回前のベクトルと同一の場合には、フィードバック処理を終了し、読み出し器7に前記部分*

8

*線形器3の出力ベクトルを出力する。前記読み出し器7では、入力した前記部分線形器3の出力ベクトルの各要素の値が大きい順に所定数選択し、それらのベクトル中での位置を抽出する。読み出し器7にはデータベース1のデータのアドレスとそれらのデータに対応する前記部分線形器3の出力ベクトル中の位置の対応関係を示すアドレステーブル11が用意されている。このテーブルにより、前記抽出した要素の前記ベクトル中の位置から前記データベース1中のアドレスを求め、そのアドレスの情報を表示器8を通して出力する。利用者が所望する情報が取り出されるまで、以上の動作が繰り返される。

【0032】一方、蓄積器44には、キーワード入力器40の出力である検索キーワードベクトルと、その検索キーワードに対して利用者が選択して読出された情報に付されていた索引キーワードベクトルの組が蓄積器44に記憶される。そして演算器45は前記蓄積器44に記憶された検索キーワードベクトルと、索引キーワードベクトルの組を基にキーワード間の相関を計算する。キーワード間の相関は条件付確率で表される。すなわち、検索キーワードのあるカテゴリーのキーワードを a 、利用者が選択した情報に付されていた索引キーワードの同一カテゴリーのキーワードを b とすると、カテゴリー s のキーワード a 、 b 間の相関 $C(a, b)$ は、

【0033】

【数2】

$$C(a, b) = P(b|a) = \frac{P(a \cap b)}{P(a)} \quad \dots (1)$$

で定義される。前記キーワード相関テーブル46に記憶されているキーワード間の相関度の更新は下記のように行なう。

(1) カテゴリー s の検索キーワードとして a がキーワード入力器40により入力され、検索キーワードベクトルとして蓄積器44に記憶される。

(2) 検索した情報に付されていた索引キーワードベクトルが蓄積器44に記憶される。

(3) 演算器45は、蓄積器44に記憶されている検索キーワードベクトルと、このベクトルに対応する索引キーワードベクトルのペアを読みだし、各カテゴリーに対応する部分ベクトルを調べ、そのカテゴリーの検索キーワードと索引キーワードを特定する。

(4) 前記(3)で特定したカテゴリー s の検索キーワードが a 、索引キーワードが b とすると、演算器45内のレジスタ $R(s, a, b)$ の値に“1”加える。ここで、 $R(s, a, b)$ はカテゴリーが s で、検索キーワードが a 、索引キーワードが b であった回数を記録するレジスタを表す。

【0034】

【数3】

(5) : カテゴリーsのP(a∩b)を式(2)により計算する。

$$P_s(a \cap b) = \frac{R(s, a, b)}{\sum_i R(s, a, i)} \quad \dots (2)$$

(6) C_s(a∩b)を計算して前記キーワード相関テーブル46を更新する。

【0035】本実施例では、利用者が検索に用いた検索キーワードと検索結果の情報に付されていた索引キーワードからキーワード間の相関度を求め、その相関度により検索キーワードを修正して検索しているので、情報を登録した人間と、検索する人間のキーワードに対する概念のズレが大きい場合にも効率的に検索ができる。なお、演算器2と演算器4は共に行列とベクトルの掛け算を実行しているので、1つの演算器でこれら2つ分の演算器の処理を実行することもできる。

【0036】また、検索器42は、フィードバック回数は、正規化器5の出力か、1回フィードバック数前の値と同じになるまで繰り返されることになっているが、これは、ある定められた回数行ってもよい。また、初めにフィードバックなしで検索し、所定数表示しても利用者の求めるデータが無い場合にフィードバックを行うようにしてもよい。

【0037】次に本発明による第2実施例としての情報検索装置について説明する。この第2実施例の構成は、第1実施例と同一であるが、構成要素の処理内容が異なるものである。

【0038】前述した第1実施例においては、検索器42は正規化器5の出力ベクトルが1回だけフィードバックをする前の正規化器5出力ベクトルと同一となった時点でフィードバックを停止し、前記読み出し器7により部分線形器3の出力ベクトルを読み出して、この部分線形器3の出力ベクトルの要素の値が大きい順に所定数選択し、そのベクトル中の位置から検索候補の情報が記憶されているデータベース中のアドレスを求めている。

【0039】しかし本実施例においては、フィードバックの回数が0、1、2、…、と1つずつ増加する毎に所定個の検索候補を前記データベース1から読み出して表示器8に表示する。表示された候補中に利用者が求める情報が含まれていない場合には、フィードバックを1回行なう。

【0040】また第1実施例のように、変換器41で修正された検索キーワードベクトルが演算器2、部分線形器3、演算器4で1回処理されると、連想記憶の1つの手法として知られる相関行列法を実行したこととなる。

【0041】いま、記憶すべきn個の登録ベクトルをx₁、(i=1, 2, …, n)とし、これら記憶ベクトルを列ベクトルとして構成される記憶行列をX=[x₁, x₂, …, x_n]、その転置行列をX^tとすれば、相関行列は、XX^tで表される。第1実施例で述べたように、pを索引キーワードベクトルとすると、X^tpの計

算結果のそれぞれの要素が、検索キーワードベクトルと各索引キーワードベクトルとの近さを表す。

【0042】前記部分線形器3は、所定の値以下の入力に対しては“0”を出力し、それ以上の場合には、入力値と同じ値を出力するものであり、検索キーワードベクトルと各索引キーワードベクトルの相関が小さいものについて“0”とするものである。フィードバック“0”での検索では、部分線形器3の出力g(X^tp)の値の大きい順に第1実施例と同様な処理により、データベース1に記憶されている情報を所定の数だけ利用者に表示する。この情報中に利用者が求める情報が含まれて無い場合には処理を進め、1回だけフィードバックする。この時、利用者に表示した情報に対応したg(X^tp)の要素は“0”とし、引き続きフィードバックにより繰り返される処理に関与しないようにする。

【0043】そして、1回だけフィードバックすると、前記部分線形器3の出力はg(X^tq)となる。ここで、qはXg(X^tp)を正規化器5で正規化したもので、前記演算器2の新たな入力ベクトルである。Xg(X^tp)は検索キーワードベクトルpと、索引キーワードベクトルの相関行列XX^tとの積からpと{x₁, x₂, …, x_n}との相関の小さいものを除いた結果となる。連想記憶の観点では、XX^tpはベクトルpにより記憶ベクトル{x₁, x₂, …, x_n}を想起した結果であるので、Xg(X^tp)は検索キーワードベクトルpにより情報登録時に付した索引キーワードベクトルを連想したものである。本実施例では、前記正規化器5は入力ベクトルに対してそのノルムが検索キーワードベクトルのノルムと等しくなるように正規化する。

【0044】また1回だけフィードバックし、フィードバック前と同様な処理が繰り返されて、g(X^tq)の値の大きい順にデータベース1に記憶されている情報を所定の数だけ利用者に表示する。表示中に利用者が求める情報が有った場合にはそこで検索処理を終了する。無かった場合には上記のフィードバック処理が繰り返される。この場合にはフィードバックを繰り返した後、正規化器5の出力が1回フィードバックする前の値と所定の範囲内で一致した時点で検索処理は終了する。

【0045】以上、前述したように本実施例では、まず、フィードバック前の処理では単純に検索キーワードに近い索引キーワードを有する情報から検索するため、検索キーワードが正しく入力された場合に好適する。しかし、キーワードが誤って入力された場合にも、フィードバックの繰り返される度に検索キーワードの連想によ

る修正が繰り返されるので有利な検索ができる。各フィールドバック毎にデータベースから読み出される検索情報の数をコントロールすることにより利用者の記憶の確かさに応じて効率的な検索を実現できる。

【0046】図9には、本発明による第3実施例としての情報検索装置の構成を示し説明する。ここで第3実施例の構成部材で、第1実施例の部材と同一のものには同じ参照符号を付しその説明を省略する。

【0047】この情報検索装置は、データベース利用者が所定のキーワードを選択入力するキーワード入力器40と、同時にそのキーワードに対する確信度を入力する確信度入力器50と、前記確信度に基づき、キーワード間の相関度を出力するキーワード相関テーブル51と、キーワード入力器40からのキーワードに基づく検索キーワードベクトルをキーワード相関テーブル51からのキーワード間の相関度に応じて修正する変換器41と、修正された検索キーワードベクトルに基づき所定の出力ベクトルを検索する検索器42と、前記検索器42の出力ベクトルの各要素の値が大きい順に所定数選択する選択器43と、選択器43に選択された出力ベクトル(アドレス)に従い、データベース1から所望のデータを読出す読み出し器7と、読み出されたデータを表示する表示部8とで構成される。

【0048】本実施例では、検索キーワードと共に、そのキーワードに対する確信度を入力し、この確信度に応じてキーワードが属するカテゴリ中の他のキーワードとの相関度を変更する。

【0049】図7に示す一例では、第1実施例と同様にカテゴリ毎のキーワード間の相関度が行列の形でキーワード相関テーブル51に記憶されている。確信度が“1”の場合(100%確信がある)には、図7(a)のように、選択したキーワードが“1”で他のキーワードは“0”、また図7(c)のように確信度が“0”の場合には選択したキーワードも他のキーワードも同じ値となる。

【0050】そして確信度が“1”と“0”の間の場合には、確信度が小さくなるに従って選択したキーワードそのものの相関度は小さくなり、それ以外のキーワードは大きくなる。この場合、選択したキーワードと概念的または感覚的に近いキーワードの方が遠いキーワードより相関度を大きく設定する。

【0051】また、図7(b)の例の場合には、行列の2列目は選択したキーワードが「夏」の場合の相関度を表すが、自分自身に対する相関度は“0.5”、「春」と「秋」に対する相関度は“0.25”、「冬」に対しては“0”であることを示している。

【0052】前記キーワード入力器40は、データベース利用者がキーボードなどにより、予め定められているキーワード中からいくつかを選択して入力する。同時にそのキーワードに対する確信度を確信度入力器50によ

り入力する。入力されたキーワードは、第1実施例と同様にカテゴリ毎にただ1個のキーワードに対応する要素のみ“1”で他の要素は全て“0”である検索キーワードベクトルとして変換器41に送られる。

【0053】前記変換器41は、キーワード入力器40が出力する検索キーワードベクトルをキーワード相関テーブル51に記憶されているキーワード間の相関度に応じて修正する。修正された検索キーワードベクトルは検索器42に入力する。前記キーワード相関テーブル51に記憶されているキーワード間の相関度は、前記確信度入力器50により入力した確信度に応じて修正する。修正された検索キーワードベクトルにより検索器42で検索処理が実行される。

【0054】本実施例の情報検索装置では検索キーワードに対する確信度により選択したキーワードのみでなく、選択しなかったキーワードの値を変化させることにより不確かなキーワードからも効率的に検索できる。

【0055】次に、本発明による第4実施例としてニューラルネットワークを利用した情報検索装置について説明する。ここで、第4実施例の構成部材で第1～第3実施例と同等の部材には同じ参照符号を付してその説明を省略する。

【0056】この第4実施例では、演算器2及び4と部分線形器3と正規化器5の処理をニューラルネットワークにより実行する。本実施例のニューラルネットワークは図5に示すような例えば、3層構造に形成される。第1の層(入力層)のユニット21a～21nは、入力した信号を第2層の全てのユニットにそのまま伝送する。第2層(中間層)と第3層(出力層)に属するユニット22a～22n及び23a～23nはニューロンユニットであり、第2層は第1実施例の演算器2と部分線形器3の処理を行ない、第3層は演算器4と正規化器5の処理を行なう。

【0057】このニューロンユニットは、図6のような構成となっている。各ユニットでは、並列に入力するそれぞれの信号値は、乗算器30a～30nにより重みメモリ31a～31nに記憶されている所定の重みパラメータと乗算される。前記重みパラメータが乗ぜられた信号値は、加算器32により全ての入力端子について加算される。加算された結果は加算器33により閾値メモリ34に記憶されている所定の値が引かれる。変換器35は、加算器33の出力に対し、テーブル36に記憶した入出力の対応関係のリストに基づいて、前記加算器33の出力を所定の線形または非線形変換する。すなわち、1個のニューロンユニットは式(3)を実行する。

【0058】

【数4】

$$z = f \left(\sum_{i=1}^N w_i y_i - \theta \right) \quad \dots (3)$$

ここで、 y_i は i 番目の入力値、 w_i は i 番目の入力の重み、 θ は閾値、 N は入力端子数、 f は線形または非線形変換関数である。

【0059】本実施例の場合に、第2層に属するニューロンユニット $22a \sim 22n$ は、 $\theta = 0$ 、 f として図3に示す部分線形関数とし、各ユニット $22a \sim 22n$ の入力端子の重みは、第1実施例で述べた索引キーワードベクトルの要素値とする。すなわち、記憶されている n 個のキーワードベクトルを x_i 、($i = 1, 2, \dots, n$) とし、これらキーワードベクトルを列ベクトルとして構成される記憶行列を $X = [x_1, x_2, \dots, x_n]$ 、その転置行列を X' とすれば、 i 番目のユニットの j 番目の端子の重み w_{ij} は、 X' の i 、 j 要素 (x_i の j 番目の要素) x_{ij} とする。第3層に属するニューロンユニット $23a \sim 23n$ では、 $\theta = 0$ 、 f としては入力が“0”以下の場合“0”で、“0”より大きい場合“1”となるステップ関数とする。このとき r 番目のユニットの s 番目の端子の重み w_{rs} は、式(1)の正規化するための行列を D とすると、行列積 DX の r 、 s 要素、すなわち、 $(DX)_{rs}$ とする。

【0060】そして、データベースに $n+1$ 番目の新たな情報を追加したときは、その新たな情報に対するキーワード情報に対応する1個のニューロンユニットを第2層に追加し、このユニットと第1層及び第3層のユニットの全てと結合し、この追加したユニットの全ての入力端子の重みと、このユニットの出力と結合する第3層の全てのユニットの入力端子の重みを与える。具体的には、新たなキーワードベクトルを x_{n+1} とすると、新たに追加された第2層のユニットと第1層の j 番目のユニットとの重み $w_{n+1,j}$ は $x_{n+1,j}$ で、第3層の r 番目のユニットとの重み $w_{r,n+1}$ は $(DX_{n+1})_r$ とすれば良い。

【0061】以上のように第4実施例では、計算処理にニューラルネットワークを用いることで、ニューラルネットワークの並列性により検索処理速度を向上することができる。また、データベースに情報を追加した場合は、単に第2層のユニットを追加し、そのユニットと第1層及び第3層のユニット間の重み係数を設定する事により時間のかかる学習など必要とせず新たなデータに対するキーワードを追加し、このデータをキーワード検索できるようになる。

【0062】次に本発明による第5実施例としての情報検索装置について説明する。前述した第1、第3実施例においては、検索キーワードベクトルをキーワード相関テーブルに記憶されているキーワード間の相関度に基づいて変換した後、検索器42で検索した。従って、検索キーワードベクトルのキーワード間の相関度に基づく変換処理は検索キーワードベクトルとキーワード間の相関度を表す行列の積により計算される。また、第1実施例の説明の部分で述べたように、検索器42の演算器2で

は、修正された検索キーワードベクトルと索引キーワード行列 X の転置行列 X' との積が計算される。

【0063】すなわち、キーワード入力値40が出力した検索キーワードベクトルに対して2個の行列が乗算されたことになる。ところで、2個の行列の積は1個の行列で表すことができるので、キーワードメモリ6にキーワード間の相関度を表す行列と X' の積の形で記憶しておくことにより変換器41を省くことができる。図10は第1実施例について変換器41を省いた場合のブロック図である。

【0064】また、第4実施例のように、検索器42の演算器2及び4、部分線形器3、正規化器5をニューラルネットワークにより構成する場合も、中間層に属するニューロンユニットの重みパラメータをキーワード間の相関度を表す行列と X' の積とすることにより変換器41を省くことができる。

【0065】本実施例では変換器41を必要としないので、検索装置の構成が簡単になり、処理時間も短縮できる。次に本発明による第6実施例としての情報検索装置について説明する。

【0066】この第6実施例では、キーワード間の相関度は利用者ごとに異なるため、個人ごとに対応するキーワード相関テーブルを構築する。図11に本実施例の一例を示す。ID入力器52により利用者のIDが入力されると、選択器54はそのIDに対応するキーワード相関テーブルをキーワード相関テーブルデータベース53から読み出し、変換器41に送る。変換器41はこのキーワード相関テーブルを用いて検索キーワードベクトルの修正を行ない、検索器42に出力する。検索器42以降の処理は第1実施例～第5実施例と同様である。なお、図11に示すID入力器52は、利用者の個人識別ができれば、どのような構成でもよい。

【0067】図12には、第7実施例としての情報検索装置の構成例の一例を示す。図11に示した装置では検索装置内に全利用者ごとのキーワード相関テーブルを記憶するデータベース53を有していた。図12に示す装置では、利用者個人が自分専用のキーワード相関テーブルを記憶したメモリーカードを持っており、検索する場合にはこのメモリーカード55からキーワード相関テーブルをインターフェース56を介して読み出し検索キーワードベクトルを修正する。また、検索装置の使用を終了した時点で、演算器45により計算した新たなキーワード相関度をメモリーカード55に記憶して、新たなキーワード相関テーブルとする。

【0068】以上のように利用者個人ごとに独自のキーワード相関テーブルを利用できるので、より効果的な検索キーワードベクトルの修正ができ、検索が効率的に実行できる。なお、本発明の上記実施態様によれば、以下如き構成が得られる。

【0069】(1) データベースを検索する情報検索装

10

20

30

40

50

置において、所定の検索キーワードを入力し、数値ベクトル化する検索キーワードベクトル化手段と、前記数値ベクトル化された検索キーワードベクトルをキーワード間の相関度に応じて修正する検索キーワードベクトル修正手段と、前記修正された検索キーワードベクトルに基づいて検索候補を出力する検索手段と、前記検索候補から、所望する情報を選択する選択手段と、前記キーワード間の相関度を修正するキーワード間相関度修正手段と、を具備することを特徴とする情報検索装置。

【0070】従って、検索キーワードをキーワード間の相関度に応じて修正され、さらにその相関度も修正される結果、確実性が不十分な検索キーワードであっても効果的に検索できる。

【0071】(2) 前記検索キーワードベクトル修正手段は、キーワードのカテゴリ毎に修正することの特徴とする前記(1)記載の情報検索装置。従って、検索キーワードがカテゴリ分けされており、キーワード間の相関が算出し易くなる。

【0072】(3) 前記検索キーワードベクトル化手段は、キーワードのカテゴリ毎にただ1つだけ選択したキーワードは1、残りを全て0であるベクトルに変換することを特徴とする前記(1)及(2)記載の情報検索装置。

【0073】従って、カテゴリ毎に検索キーワードを1つだけ選択し、残りを全て0とするため、計算処理が簡単になり処理時間が短くなる。

(4) 前記キーワード間相関度修正手段が、前記検索キーワードを表す数値ベクトルと、前記選択した情報に付されている索引キーワードを表す数値ベクトルの複数の組により算出する手段により構成されていることを特徴とする前記(1)乃至(3)記載の情報検索装置。

【0074】従って、検索キーワードと利用者が実際に選択した情報に付されていた索引キーワードにより相関度が求められ、キーワード間の相関度の信頼性が高くなる。

(5) データベースを検索する情報検索装置において、所定の検索キーワードを入力し数値ベクトル化する検索キーワードベクトル化手段と、前記数値ベクトル化された検索キーワードベクトルを、キーワードのカテゴリ毎にキーワード間の相関度に応じて修正する検索キーワードベクトル修正手段と、前記修正された検索キーワードベクトルに基づいて、検索候補を出力する検索手段と、前記検索候補から、所望の情報を選擇する選擇手段と、選擇された所望情報に数値ベクトルの形で付されている索引キーワードベクトルを前記検索キーワードベクトルと共に記憶するキーワードベクトル記憶手段と、前記キーワードベクトル記憶手段に記憶されている複数の検索及び索引キーワードベクトルの組から前記キーワード間の相関度を算出するキーワード相関度算出手段と、前記キーワード相関度算出手段により算出された値

により前記キーワード間の相関度を修正するキーワード間相関度修正手段と、を具備することを特徴とする情報検索装置。

【0075】従って、検索キーワードをキーワード間の相関度に応じて修正する機能と、キーワード間の相関度を実際の検索結果から求める手段を有するので、不十分な検索キーワードからも効果的に検索できる。

【0076】(6) 前記検索器は検索候補をベクトルとして出力し、そのベクトルから検索候補の情報の存在するアドレスを指定する手段を、さらに具備することを特徴とする前記(5)記載の情報検索装置。

【0077】従って、検索候補をベクトルの形で同時に求めることができ、検索の効率化が実現できる。

(7) データベースを検索する情報検索装置において、利用者を特定する手段と、利用者毎に前記キーワード相関テーブルと、を具備することを特徴とする前記(1)乃至(6)記載の情報検索装置。

【0078】従って、検索装置内に利用者毎のキーワード間の相関度を保持し、検索時に利用者を特定してその利用者のキーワード間の相関度により検索キーワードを修正するため、精度の高い修正ができる。

【0079】(8) データベースを検索する情報検索装置において、検索時に利用者が自分専用のキーワード相関テーブルを与える手段を、さらに具備することを特徴とする(7)記載の情報検索装置。

【0080】従って、検索装置の外部に利用者毎のキーワード相関度テーブルを有し、検索時に利用者固有のキーワード相関度テーブルを読み取り、それにより検索キーワードを修正するので精度の高い修正ができ、かつ検索装置の規模を小さくできる。また、種々の検索装置において同一のキーワード相関度テーブルを利用できる。

【0081】(9) データベースを検索する情報検索装置において、前記利用者個人の専用のキーワード相関テーブルを情報検索装置の外部の記憶媒体に保持していることを特徴とする前記(7)及び(8)記載の情報検索装置。従って、検索装置の外部に利用者毎のキーワード相関度テーブルを有し、検索時に利用者固有のキーワード相関度テーブルを読み込んで、それにより検索キーワードを修正するので精度の高い修正が簡単にでき、かつ検索装置の規模を小さくできる。また、種々の検索装置において同一のキーワード相関度テーブルを利用できる。

【0082】(10) データベースを検索する情報検索装置において、所定の検索キーワードデータを入力し数値ベクトル化する検索キーワードベクトル化手段と、その検索キーワードの確信度を入力する手段と、前記数値ベクトル化された検索キーワードベクトルを前記確信度に応じて修正する検索キーワードベクトル修正手段と、前記修正された検索キーワードベクトルを入力し、検索候補を出力する検索手段と、前記検索候補から、所

望の情報を選択する選択手段と、具備することを特徴とする情報検索装置。

【0083】従って、確信度によりキーワード相関度テーブルが修正され、利用者の記憶の状況を検索に反映させることができる。

(11) 前記検索キーワードベクトル修正手段は、検索キーワードのカテゴリ毎に修正し、確信度が1の場合には未修正とし、確信度が0の場合にはそのキーワードが属するカテゴリの全てのキーワードに同じ数値となるように修正し、1と0の間の場合には選択したキーワードについては値を小さくし、それ以外のキーワードについては値を大きくすることを特徴とする前記(10)記載の情報検索装置。

【0084】従って、確信度によりキーワード相関度テーブルが修正され、利用者の記憶の状況を検索に反映させることができる。

(12) 前記検索手段が連想過程を1乃至複数回繰り返しながら行うことにより検索を実行することを特徴とする前記(1)乃至(11)記載の情報検索装置。

【0085】従って、検索処理が連想過程を1乃至複数回繰り返しながら行なうことにより検索を実行するので、利用者が入力した検索キーワードの正しさの度合に応じて効率的な検索が実現できる。

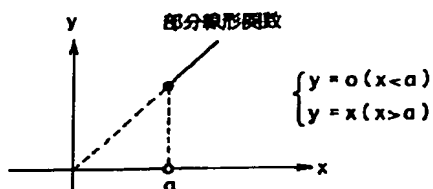
【0086】(13) 前記検索手段がニューラルネットワークであることを特徴とする(12)記載の情報検索装置。従って、ニューラルネットワークにより処理するので並列的に実行でき高速の処理が可能である。

【0087】(14) 前記検索手段が前記検索キーワードベクトル修正手段を含んで処理することを特徴とする前記(1)乃至(13)記載の情報検索装置。従って、検索キーワードの修正処理を検索処理に含めることができるので、本発明の検索装置の規模を小さくすることができる。

【0088】(15) 前記検索手段が単一の演算器により構成されることを特徴とする前記(1)乃至(14)記載の情報検索装置。従って、検索手段における演算を単一の演算器により実行するので、本発明の検索装置の規模を小さくすることができる。

【0089】

【図3】



*【発明の効果】以上詳述したように本発明によれば、指示された検索キーワードを予め定められたキーワード間の相関度に応じて検索に好適するように修正し、データベースから所望する情報を取り出す情報検索装置を提供することができる。

【図面の簡単な説明】

【図1】本発明の情報検索装置な具体的な構成例を示す図である。

【図2】カテゴリの一例を示す図である。

10 【図3】部分線形器における部分線形関数の一例を示す図である。

【図4】線形変換器の出力ベクトルとデータベースとの関係を示す図である。

【図5】本発明の第4実施例としての情報検索装置に用いられるニューラルネットワークの一例を示す図である。

【図6】本実施例に用いるニューロンユニットの構成例を示す図である。

【図7】キーワード相関テーブルに記憶されているカテゴリ毎のキーワード間の相関度の行列の一例を示す図である。

20 【図8】本発明による第1実施例としての情報検索装置を含むファイル検索システムの概略的な構成を示す図である。

【図9】本発明による第3実施例としての情報検索装置の構成を示す図である。

【図10】第5実施例として、第1実施例から変換器を省いた情報検索装置の構成を示す図である。

30 【図11】第6実施例としての情報検索装置の構成例の一例を示す図である。

【図12】第7実施例としての情報検索装置の構成例の一例を示す図である。

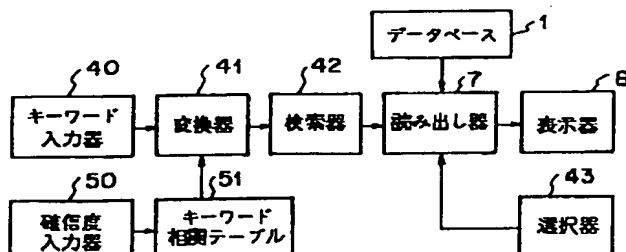
【符号の説明】

1…データベース、2…演算器、3…部分線形器、4…第2の演算器、5…正規化器、6…キーワードメモリ、7…読み出し器、8…表示器、9…、10…コントローラ、11…アドレステーブル、40…キーワード入力器、41…変換器、42…検索器、43…選択器、44…変換器、45…検索器、46…選択器。

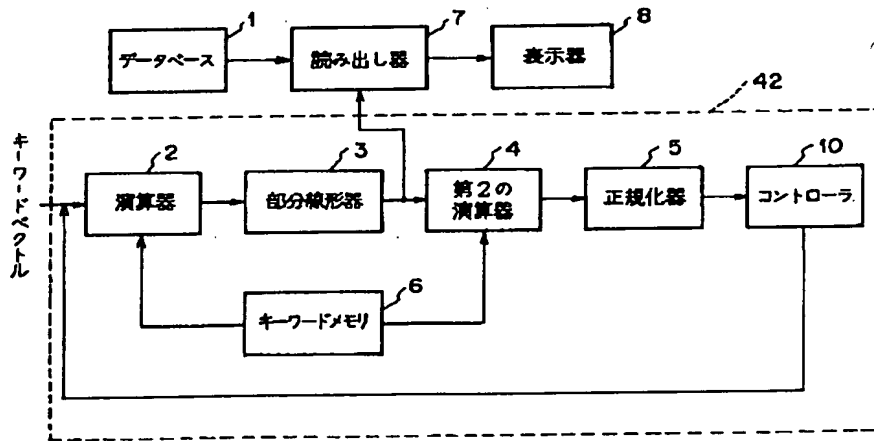
*

40

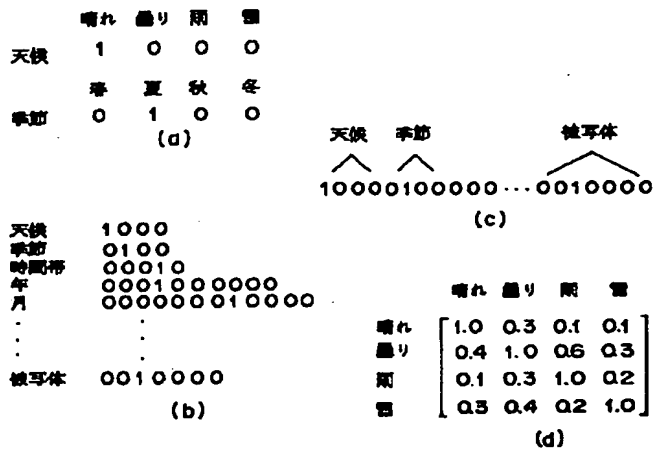
【図9】



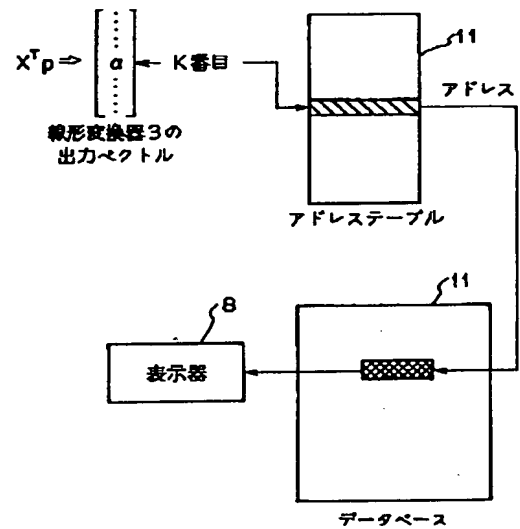
【図1】



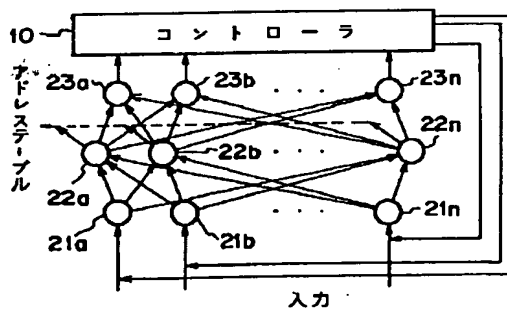
【図2】



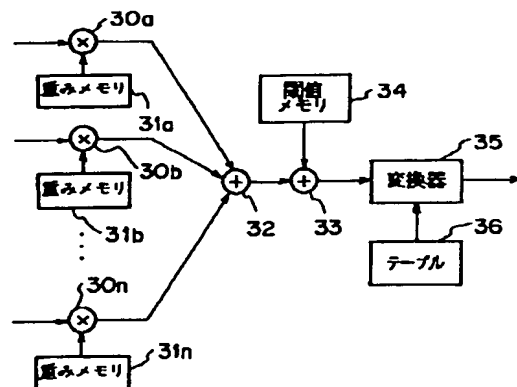
【図4】



【図5】



【図6】



【図7】

	春	夏	秋	冬
春	1.0	0.0	0.0	0.0
夏	0.0	1.0	0.0	0.0
秋	0.0	0.0	1.0	0.0
冬	0.0	0.0	0.0	1.0

(a) 確信度 1.0

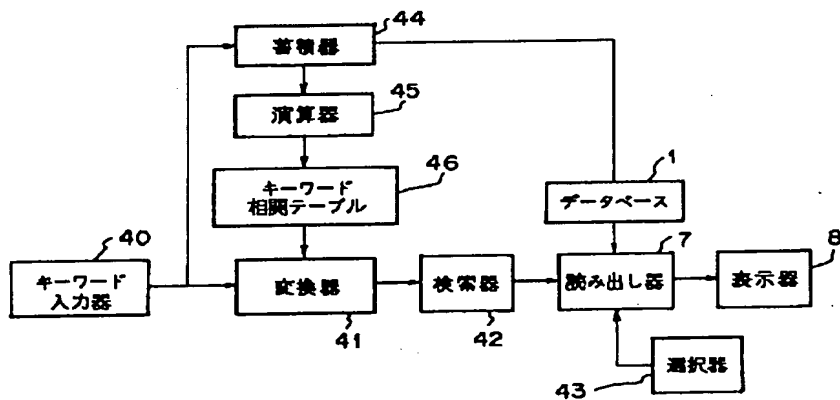
	春	夏	秋	冬
春	0.5	0.25	0.2	0.25
夏	0.15	0.5	0.15	0.0
秋	0.2	0.25	0.5	0.25
冬	0.15	0.0	0.15	0.5

(b) 確信度 0.5

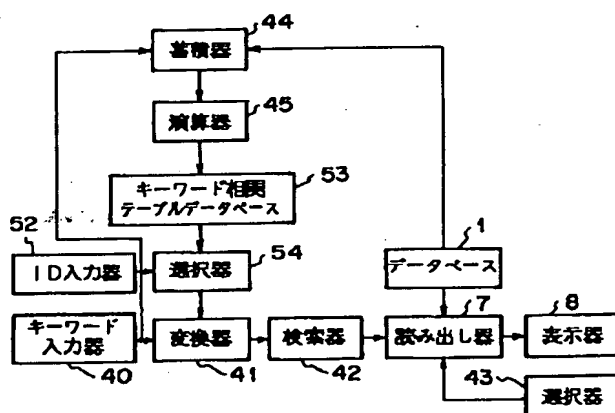
	春	夏	秋	冬
春	0.25	0.25	0.25	0.25
夏	0.25	0.25	0.25	0.25
秋	0.25	0.25	0.25	0.25
冬	0.25	0.25	0.25	0.25

(c) 確信度 0.0

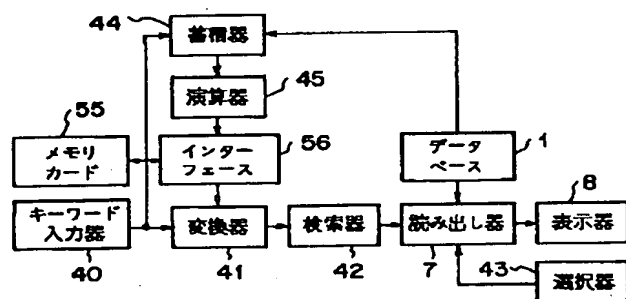
【図8】



【図11】



【図12】



【図10】

